(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 July 2001 (26.07.2001)

PCT

(10) International Publication Number
WO 01/53312 A1

(51) International Patent Classification[7]: C07H 21/04, C12N 15/11, 15/63, 15/70, 15/82, 15/85, C07K 14/00

(21) International Application Number: PCT/US00/34263

(22) International Filing Date:
26 December 2000 (26.12.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

09/488,725	21 January 2000 (21.01.2000)	US
09/552,317	25 April 2000 (25.04.2000)	US
09/598,042	9 July 2000 (09.07.2000)	US
09/620,312	19 July 2000 (19.07.2000)	US
09/653,450	3 August 2000 (03.08.2000)	US
09/662,191	14 September 2000 (14.09.2000)	US
09/693,036	19 October 2000 (19.10.2000)	US
09/727,344	29 November 2000 (29.11.2000)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	09/488,725 (CIP)
Filed on	21 January 2000 (21.01.2000)
US	09/552,317 (CIP)
Filed on	25 April 2000 (25.04.2000)
US	09/598,042 (CIP)
Filed on	9 July 2000 (09.07.2000)
US	09/620,312 (CIP)
Filed on	19 July 2000 (19.07.2000)
US	09/653,450 (CIP)
Filed on	3 August 2000 (03.08.2000)
US	09/662,191 (CIP)
Filed on	14 September 2000 (14.09.2000)
US	09/693,036 (CIP)
Filed on	19 October 2000 (19.10.2000)
US	09/727,344 (CIP)
Filed on	29 November 2000 (29.11.2000)

(71) Applicant *(for all designated States except US)*: **HYSEQ, INC.** [US/US]; 670 Almanor Avenue, Sunnyvale, CA 94086 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **LIU, Chenghua** [CN/US]; 1125 Ranchero Way #14, San Jose, CA 95117 (US). **ASUNDI, Vinod** [US/US]; 709 Foster City Boulevard, Foster City, CA 94404 (US). **CHEN, Rui-hong** [US/US]; 1031 Flying Fish Street, Foster City, CA 94404 (US). **MA, Yunqing** [CN/US]; 280 W. California Avenue #206, Sunnyvale, CA 94086 (US). **QIAN, Xiaohong, B.** [CN/US]; 3662 Tumble Way, San Jose, CA 95132 (US). **REN, Feiyan** [US/US]; 7703 Oak Meadow Court, Cupertino, CA 95014 (US). **WANG, Dunrui** [CN/US]; 932 La Palma, Milpitas, CA 95035 (US). **WANG, Jian-Rui** [CN/US]; 744 Stendhal Lane, Cupertino, CA 95014 (US). **WANG, Zhiwei** [CN/US]; 836 Alturas Avenue, B36, Sunnyvale, CA 94085 (US). **WEHRMAN, Tom** [US/US]; 3210 CCSR Mol Pharm, 269 W. Campus Drive, Stanford, CA 94305 (US). **XU, Chongjun** [CN/US]; 4918 Manitoba Drive, San Jose, CA 95130 (US). **XUE, Aidong, J.** [CN/US]; 1621 S. Mary Avenue, Sunnyvale, CA 94087 (US). **YANG, Yonghong** [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **ZHANG, Jie** [CN/US]; 4930 Poplar Terrace, Campbell, CA 95008 (US). **ZHAO, Qing, A.** [CN/US]; 1556 Kooser Road, San Jose, CA 95118 (US). **ZHOU, Ping** [CN/US]; 1461 Japaul Lane, San Jose, CA 95132 (US). **GOODRICH, Ryle** [US/US]; 4896 Sandy Lane, San Jose, CA 95124 (US). **DRMANAC, Radoje, T.** [YU/US]; 850 East Greenwich Place, Palo Alto, CA 94303 (US).

(74) Agent: **ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris, Glovsky, and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
with international search report
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

[Continued on next page]

(54) Title: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

(57) Abstract: The present invention provides novel nucleic acids, novel polypeptide sequences encoded by these nucleic acids and uses thereof.

WO 01/53312 A1

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

SEQ ID NO:	PFAM NAME	DESCRIPTION	p-value	PFAM SCORE
	t			
429	zf-C3HC4	Zinc finger, C3HC4 type (RING finger)	8.6e-11	39.2
431	DEAD	DEAD/DEAH box helicase	1e-66	214.0
432	SH3	SH3 domain	3.4e-16	67.2
433	GTP_CDC	Cell division protein	2.1e-114	393.5
436	Collagen	Collagen triple helix repeat (20 copies)	4.6e-194	658.1
438	Ricin_B_lectin	Similarity to lectin domain of ricin b	0.0085	10.5
441	Alpha_adaptin_C	Alpha adaptin carboxyl-terminal domai	1.2e-256	866.0
442	Alpha_adaptin_C	Alpha adaptin carboxyl-terminal domai	1.8e-235	795.7
443	PDZ	PDZ domain (Also known as DHR or GLGF).	1.9e-65	230.9
445	LON	ATP-dependent protease La (LON) domain	0.00012	-17.1
446	ig	Immunoglobulin domain	0.00011	20.1
451	sushi	Sushi domain (SCR repeat)	1.4e-18	75.2
452	fn3	Fibronectin type III domain	1.5e-06	35.2
454	pyridoxal_dec	Pyridoxal-dependent decarboxylase conse	8.3e-14	50.3
456	kinesin	Kinesin motor domain	4.9e-217	734.4
457	neur_chan	Neurotransmitter-gated ion-channel	1e-175	597.1
458	Josephin	Josephin	0.0002	18.7
468	bZIP	bZIP transcription factor	1.7e-07	31.8
470	NTP_transferase	Nucleotidyl transferase	6.3e-06	-26.3
471	WD40	WD domain, G-beta repeat	2e-28	107.9
473	LIM	LIM domain containing proteins	0.00021	20.7
477	zf-RanBP	Zn-finger in Ran binding protein and others.	0.028	21.0
479	WD40	WD domain, G-beta repeat	6.5e-18	73.0
480	KRAB	KRAB box	1e-31	118.8
481	ArfGap	Putative GTP-ase activating protein for Arf	8.4e-66	232.0
485	SH2	Src homology domain 2	0.011	11.4
486	C1q	C1q domain	4.3e-74	259.6
487	dsrm	Double-stranded RNA binding motif	1.1e-47	171.9
489	zf-C2H2	Zinc finger, C2H2 type	4.8e-153	521.9
490	Alpha_adaptin_C	Alpha adaptin carboxyl-terminal domai	3.4e-222	751.6
492	SKI	Shikimate kinase	1.2e-10	48.8
497	ENV_polyprotein	ENV polyprotein (coat polyprotein)	2.6e-22	77.6
498	abhydrolase_2	Phospholipase/Carboxylesterase	0.041	-48.1
500	rrm	RNA recognition motif.	5.4e-34	126.4
501	WW	WW domain	4.6e-18	73.4
502	ig	Immunoglobulin domain	1.1e-10	39.5
504	abhydrolase	alpha/beta hydrolase fold	0.045	-3.6
505	vwa	von Willebrand factor type A domain	7.1e-62	219.0
508	Na_K_ATPase_C	Na+/K+ ATPase C-terminus	2.3e-145	496.3
509	Exonuclease	Exonuclease	1.3e-56	201.5
510	Glycos_transf_1	Glycosyl transferases group 1	2.9e-06	27.0
511	Glycos_transf_1	Glycosyl transferases group 1	2.9e-06	27.0
512	Glycos_transf_1	Glycosyl transferases group 1	1.9e-09	38.5
514	pro_isomerase	Cyclophilin type peptidyl-prolyl cis-tr	1.8e-63	221.4

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Priority docket number corresponding SEQ ID NO: in priority application	SEQ ID NO:in U.S.S.N. 09/488,725
432	2218	4004	5790	784CIP2B_106	3417
433	2219	4005	5791	784CIP2B_107	3418
434	2220	4006	5792	784CIP2B_108	3442
435	2221	4007	5793	784CIP2B_109	3442
436	2222	4008	5794	784CIP2B_110	3444
437	2223	4009	5795	784CIP2B_111	3855
438	2224	4010	5796	784CIP2B_112	3863
439	2225	4011	5797	784CIP2B_113	4090
440	2226	4012	5798	784CIP2B_114	4105
441	2227	4013	5799	784CIP2B_115	4142
442	2228	4014	5800	784CIP2B_116	4142
443	2229	4015	5801	784CIP2B_117	4149
444	2230	4016	5802	784CIP2B_118	4196
445	2231	4017	5803	784CIP2B_119	4202
446	2232	4018	5804	784CIP2B_120	4274
447	2233	4019	5805	784CIP2B_121	4304
448	2234	4020	5806	784CIP2B_122	4306
449	2235	4021	5807	784CIP2B_123	4311
450	2236	4022	5808	784CIP2B_124	4321
451	2237	4023	5809	784CIP2B_125	4323
452	2238	4024	5810	784CIP2B_126	4332
453	2239	4025	5811	784CIP2B_127	4488
454	2240	4026	5812	784CIP2B_128	4588
455	2241	4027	5813	784CIP2B_129	5569
456	2242	4028	5814	784CIP2B_130	5573
457	2243	4029	5815	784CIP2B_131	5577
458	2244	4030	5816	784CIP2B_132	5579
459	2245	4031	5817	784CIP2B_133	5582
460	2246	4032	5818	784CIP2B_134	5583
461	2247	4033	5819	784CIP2B_135	5584
462	2248	4034	5820	784CIP2B_136	5585
463	2249	4035	5821	784CIP2B_137	5591
464	2250	4036	5822	784CIP2B_138	5593
465	2251	4037	5823	784CIP2B_139	5594
466	2252	4038	5824	784CIP2B_140	5594
467	2253	4039	5825	784CIP2B_141	5598
468	2254	4040	5826	784CIP2B_142	5602
469	2255	4041	5827	784CIP2B_143	5605
470	2256	4042	5828	784CIP2B_144	5608
471	2257	4043	5829	784CIP2B_145	5617
472	2258	4044	5830	784CIP2B_146	5620
473	2259	4045	5831	784CIP2B_147	5622
474	2260	4046	5832	784CIP2B_148	5623
475	2261	4047	5833	784CIP2B_149	5624
476	2262	4048	5834	784CIP2B_150	5625
477	2263	4049	5835	784CIP2B_151	5627
478	2264	4050	5836	784CIP2B_152	5628
479	2265	4051	5837	784CIP2B_153	5630
480	2266	4052	5838	784CIP2B_154	5632
481	2267	4053	5839	784CIP2B_155	5640
482	2268	4054	5840	784CIP2B_156	5641
483	2269	4055	5841	784CIP2B_157	5643
484	2270	4056	5842	784CIP2B_158	5647
485	2271	4057	5843	784CIP2B_159	5649
486	2272	4058	5844	784CIP2B_160	5658
487	2273	4059	5845	784CIP2B_161	5659
488	2274	4060	5846	784CIP2B_162	5667
489	2275	4061	5847	784CIP2B_163	5672
490	2276	4062	5848	784CIP2B_164	5674
491	2277	4063	5849	784CIP2B_165	5678
492	2278	4064	5850	784CIP2B_166	5680
493	2279	4065	5851	784CIP2B_167	5684

SEQ ID NO:	Predicted beginning nucleotide location corresponding to first amino acid residue of amino acid sequence	Predicted end nucleotide location corresponding to first amino acid residue of amino acid sequence	Amino acid segment containing signal peptide (A=Alanine, C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop Codon, /=possible nucleotide deletion, \=possible nucleotide insertion)
			LMMLLRPLLVKKIACGLGKSDRFKSIYAALYFFPILTVLQAVGG GLLYYAFPYIILVLSLVTLAVYMSASEIENCYDLLVRKKRLIVL FSHWLLHAYGIISISRVDKLEQDLPLLALVPTPALFYLFTAKFT EPSRILSECANGH
5808	2	433	SLPDSGVVEYLSNGGVADNHKDFGSLRYNECLMNFSCNGKNGSS EGRITHGFQLKSAYENNLMPYTNYTFDFKGVIDYIFYSKTHMNV LGVLGPLDPQWLVENNITGCPHPHIPSDHFSLLTQLELHPPLLP LVNGVHLPNRR
5809	464	2422	ILVPGFQGILHPGVYCALQSQHQAQELVADIDECEVSGLCRHGG RCVNTHGSFECYCMDGYLPRNGPEFFHPTTDATSCTEIDCGTPP EVPDGYIIGNYTSSLGSQVRYACREGFFSVPEDTVSSCTGLGTW ESPKLHCQEINCGNPPEMRHAILVGNHSSRLGGVARYVCQEGFE SPGGKITSVCTEKGTWRESTLTCTEILTKINDVSLFNDTCVRWQ INSRRINPKISYVISIKGQRLDPMESVRETVNLTTDSRTPEVC LALYPGTNYTVNISTAPPRRSMPAVIGFQTAEVDLLEDDGSFNI SIFNETCLKLNRRSRKVGSEHMYQFTVLGQRWYLANFSHATSPN FTTREQVPVVCLDLYPTTDYTVNVTLLRSPKRHSVQITIATPPA VKQTISNISGFNETCLRWRSIKTADMEEMYLFHIIWGQRWYQKEF AQEMTFNISSSSRDPEVCLDLRPGTNYNVSLRALSSELPVVISL TTQITEPPLPEVEFFTVHRGPLPRLRLRKAKEKNGPISSYQVLV LPLALQSTFSCDSEGASSFFSNASDADGYVAAELLAKDVPDDAM EIPIGDRLYYGEYYNAPLKRGSDYCIILRITSEWNKVRRHSCAV WAQVKDSSLMLLQMAGVGLGSLAVVIILTFLSFSAV
5810	3	1641	KVFGTHKDHEVSTLDTAISAVKVQLAEFLENLQEKSLRIEAFVS EIESFFNTIEENCSKNEKRLEEQNEEMMKKVLAQYDEKAQSFEE VKKKKMEFLHEQMVHFLQSMDTAKDTLETIVREAEELDEAVFLT SFEEINERLLSAMESTASLEKMPAAFSLFEHYDDSSARSDQMLK QVAVPQPPRLEPQEPNSATSTTIAVYWSMNKEDVIDSFQVYCME EPQDDQEVNELVEEYRLTVKESYCIFEDLEPDRCYQVWVMAVNP TGCSLPSERAIFRTAPSTPVIRAEDCTVCWNTATIRWRPTTPEA TETYTLEYCRQHSPEGEGLRSPSGIKGLQLKVNLQPNDNYFFYV RAINAFGTSEQSEAALISTRGTRFLLLRETAHPALIISSSGTVI SFGERRRLTEIPSVLGEELPSCGQHYWETTVTDCPAYRLGICSS SAVQAGALGQGETSNYMHCSEPQRYTFFYSGIVSDVHVTERPAR VGILLDYNNQRLIFINAESEQLLFIIRHRFNEGVHPAPALEKPG KCTLHLGIEPPDSVRHK
5811	1918	851	AAALADPLPEDKWSAEKRRPLKSSLGYEITFSLLNPDPKSHDVY WDIEGAVRRYVQPFLNALGAAGNFSVDSQILYYAMLGVNPRFDS ASSSYYLDMHSLPHVINPVESRLGSSAASLYPVLNFLLYVPELA HSPLYIQDKDGAPVATNAPHSPRWGGIMVYNVDSKTYNASVLPV RVRVDMVRVMEVFLAQLRLLFGIAQPQLPPKCLLSGPTSEGLMT WELDRLLWARSVENLATATTTLTSLAQLLGKISNIVIKDDVASE VYKAVAAVQKSAEELASGHLASAFVASQEAVTSSRLAFFDPSLL HLLYPPDDQKFAIYIPLFLPMAVPILLSLVKIFLETRKSWRKPE KTD
5812	5204	2744	GGRQRCQRGRSCGAREEEVEPGTARPPPAASAMDASLEKIADPI LAEMGKNLKEAVKMLEDSQRRTEEENGKKLISGDIPGPLQGSGQ DMVSILQLVQNLMHGDEDEEPQSPRIQNIGEQGHMALLGESLGA YISTLDKEKLRKLTTRILSDTTLWLCRIFRYKNGCAYFHEEERE GLAKICRLAIHSRYEDFVVDGFNVLYNKKPVIYLSAAARPGLGQ YLCNQLGLPFPCLCRVPCNTVFGSQHQMDVAFLEKLIKDDIERG RLPLLLVANAGTAAVGHTDKIGRLKELCEQYGIWLHVEGVNLAT LALGYVSSSVLAAAKCDSMTMTPGPWLGLPAVPAVTLYKHDDPA LTLVAGLTSNKPTDKLRALPLWLSLQYLGLDGFVERIKHACQLS QRLQESLKKVNYIKILVEDELSSPVVVFRFFQELPGSDPVFKAV PVPNMTPSGVGRERHSCDALNRWLGEQLKQLVPASGLTVMDLEA EGTCLRFSPLMTAAVLGTRGEDVDQLVACIESKLPVLCCTLQLR EEFKQEVEATAGLLYVDDPNWSGIGVVRYEHANDDKSSLKSYPQ GENIHAGLLKKLNELESDLTFKIGPEYKSMKSCLYVGMASDNVR AAELVETIAATAREIEDNSRLLENMTEVVRKGIQEAQVELQKAS EERLLEEGVLRQIPVVGSVLNWFSPVQALQKGRTFNLTAGSLES

SEQ ID NO:	Predicted beginning nucleotide location corresponding to first amino acid residue of amino acid sequence	Predicted end nucleotide location corresponding to first amino acid residue of amino acid sequence	Amino acid segment containing signal peptide (A=Alanine, C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop Codon, /=possible nucleotide deletion, \=possible nucleotide insertion)
			TEPIYVYKAQGAGVTLPPTPSGSRTKQRLPGQKPFKRSLRGSDA LSETSSVSHIEDLEKVERLSSGPEQITLEASSTEGHPGADSPQH TDQTEAFQKGVPHPEDDHSQVEGPESLR
5813	2936	699	HRDGVSGSLERPLTDRGRTGAFAQQRGKMATAGGGSGADPGSRG LLRLLSFCVLLAGLCRGNSVERKIYIPLNKTAPCVRLLNATEQI GCQSSISGDTGVIHVVEKEEDLQWVLTDGPNPPYMVLLESKHPT RDLMEKLKGRTSRIAGLAVSLTKPSPASGFSPSVQCPNDGFGVY SNSYGPEFAHCREIQWNSLGNGLAYEDFSFPIPLLEDENETKVI KQCYQDHNLSQNGSAPTFPLCAMQLFSHMAWLSFSTAT\CMRRS SIQSTFSINPKIVCDPLSDYNVWSMLKPINTTGTLKPDDRVVVA ATRLDSRSFFWNV\APGAESAVASFVTQLAAABALQKAPDVTTL PRNVMFVPFQGETFDYTGSSRMVYDMEKGKFPVQLENVDSPVEL GQVALRTSLELNMHTDPVSQKNESVRNQVEDLLATLEKSGAGVP AVILRRPNQSQPLPPSSLQRFLRARNISGVVLADHSGAFHNKYY QSIYDTAENINVSYPEWLEPLKE/ETWNFG*QDTAKALADVATV LGRALYELAGGTNFSDTVQADFQTVTRLLYG\FLIKANNSWFQS ILQGRDLRSYLG*RGLFQH\YIAV\SSPTNTIYV/VLQYALANL TGTVVNLTREQCQDPSKVPSENKDLYEYSWVQGPLHSNETDRLP RCVRSTARLARALSPAFELSQWSSTEVSTWTESRWKDIRARIFL IASKELELITLTVGFGILIFSLIVTYCINAKADVLFIAPREPGA VSY
5814	8500	432	ALKCRPRRVLAILVGFVQPDRMAEEGAVAVCVRVRPLNSREESL GETAQVYWKTHNNVIYPVDGSKSFNFDRVLHGNETPKNVYEA\I AAPIIDSAIQGYNGTIFA\YGQT\ASGKTYTMMGSEDHLGVIPQ GQFHGHFSQKI*EVFLDREFLLRVSYMEIYNETITDLLCGTQKM KPLIIREDVNRNVYVADLTEEVVYTSEMALKWITKGERSRHYGE TKMNQRSSRSHTIFRMILESRERGERPSNCEGSVKVSHLNLVDLA GSERAAQTGAAGVRLKEGCNINRSLFILGQVIKKLSDGQVGGFI NYRDSKLTRILQNSLGGNPKTRIICTITPVSFDETLTALQFAST AKYMKNTPYVNEVSTDEALLKRYREEIMDLKKQLEEVSLETRAQ AMEKDQLAQLLEKKDLFQKVQNEKIENLTRMLVTSSSLTLQQEL KAKRKRRVTWCLGKINKMKNSNYADQFNIPTNITTKTHKLSINL LREIDESVCSESDVFSNTLDTLSEIEWNPATKLLNQENIESELN SLRADYDNLVLDYEQLRTEKEEMELKLKEKNDLDSFEALERKTK KDQEMQLIHEISNLKNLVKHREVYNQDLENELSSKVELLREKED QIKKLQEYIDSQKLENIKMDLSYSLESIEDPKQMKQTLFDAETV ALDAKRESAFLRSENLELKEEMKELATTYKQMENDIQLYQSQLE AKKKMQVDLEKELQSAFNEITKLTSLIDGKVPKDLLCNLELEGK ITDLQKELNKEVEENEALREEVILLSELKSLPSEVERLRKEIQD KSEELHIITSEKDKLFSEVVHKESRVQGLLEEIGKTKDDLATTQ SNYKSTDQEFQNFKTLHMDFEQKYKMVLESNERMNQEIVNLSKE AQKFDSSLGALKTELSYKTQELQEKTREVQERLNEMEQLKEQLE NRDSPLQTVEREKTLITEKLQQTLEEVKTLTQEKDDLKQLQESL QIERDQLKSDIHDTVNMNIDTQEQLRNALESLKQHQETINTLKS KISEEVSRNLHMEENTGETKDEFQQKMVGIDKKQDLEAKNTQTL TADVKDNEIISQQRKIFSLIQEKNELQQMLESVIAEKEQLKTDL KENIEMTIENQEELRLLGDELKKQQEIVAQEKNHAIKKEGELSR TCDRLAEVEEKLKEKSQQLQEKQQQLLNVQEEMSEMQKKINEIE NLKNELKNKELTLEHMETERLELAQKLNENYEEVKSITKERKVL KELQKSPETERDHLRGYIREIEATGLQTKEELKIAHIHLKEHQE TIDELRRSVSEKTAQIINTQDLEKSHTKLQEEIPVLHEEQELLP NVKKVSETQETMNELELLTEQSTTKDSTTLARIEMERLRLNEKF QESQEEIKSLTKERDNLKTIKEALEVKHDQLKEHIRETLAKIQE SQSKQEQSLNMKEKDNETTKIVSEMEQPKPKDSALLRIEIEMLG LSKRLQESHDEMKSVAKEKDDLQRLQEVLQSESDQLKENIKEIV AKHLETEEELKVAHCCLKEQEETINELRVNLSEKETEISTIQKQ LEAINDKLQNKIQEIYEKEEQLNIKQISEVQRKVNELKQFKEHR KAKDSALQSIESKMLELTNRLQESQEEIQIMIKEKEEMKRVQEA LQIERDQLKENTKEIVAKMKESQEKEYQFLKMTAVNETQEKMCE IEHLKEQFETQKLNLENIETENIRLTQILHENLEEMRSVTKERD DLRSVEETLKVERDQLKENLRETITRDLEKQEELKIVHMHLKEH

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO:1-1786 and 3573-5358, a mature protein coding portion of SEQ ID NO:1-1786 and 3573-5358, an active domain of SEQ ID NO:1-1786 and 3573-5358, and complementary sequences thereof.

2. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide hybridizes to the polynucleotide of claim 1 under stringent hybridization conditions.

3. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide has greater than about 90% sequence identity with the polynucleotide of claim 1.

4. The polynucleotide of claim 1 wherein said polynucleotide is DNA.

5. An isolated polynucleotide of claim 1 wherein said polynucleotide comprises the complementary sequences.

6. A vector comprising the polynucleotide of claim 1.

7. An expression vector comprising the polynucleotide of claim 1.

8. A host cell genetically engineered to comprise the polynucleotide of claim 1.

9. A host cell genetically engineered to comprise the polynucleotide of claim 1 operatively associated with a regulatory sequence that modulates expression of the polynucleotide in the host cell.

10. An isolated polypeptide, wherein the polypeptide is selected from the group consisting of:

(a) a polypeptide encoded by any one of the polynucleotides of claim 1; and

(b) a polypeptide encoded by a polynucleotide hybridizing under stringent conditions with any one of SEQ ID NO:1-1786 and 3573-5358.

11. A composition comprising the polypeptide of claim 10 and a carrier.

12. An antibody directed against the polypeptide of claim 10.

13. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polynucleotide of claim 1 for a period sufficient to form the complex; and

b) detecting the complex, so that if a complex is detected, the polynucleotide of claim 1 is detected.

14. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample under stringent hybridization conditions with nucleic acid primers that anneal to the polynucleotide of claim 1 under such conditions;

b) amplifying a product comprising at least a portion of the polynucleotide of claim 1; and

c) detecting said product and thereby the polynucleotide of claim 1 in the sample.

15. The method of claim 14, wherein the polynucleotide is an RNA molecule and the method further comprises reverse transcribing an annealed RNA molecule into a cDNA polynucleotide.

16. A method for detecting the polypeptide of claim 10 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex; and

b) detecting formation of the complex, so that if a complex formation is detected, the polypeptide of claim 10 is detected.

17. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10 under conditions sufficient to form a polypeptide/compound complex; and

b) detecting the complex, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

18. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10, in a cell, under conditions sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and

b) detecting the complex by detecting reporter gene sequence expression, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

19. A method of producing the polypeptide of claim 10, comprising,

a) culturing a host cell comprising a polynucleotide sequence selected from the group consisting of a polynucleotide sequence of SEQ ID NO:1-1786 and 3573-5358, a mature protein coding portion of SEQ ID NO:1-1786 and 3573-5358, an active domain of SEQ ID NO:1-1786 and 3573-5358, complementary sequences thereof and a polynucleotide sequence hybridizing under stringent conditions to SEQ ID NO:1-1786 and 3573-5358, under conditions sufficient to express the polypeptide in said cell; and

b) isolating the polypeptide from the cell culture or cells of step (a).

20. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of any one of the polypeptides SEQ ID NO:1787 -3572 and 5359-7144, the mature protein portion thereof, or the active domain thereof.

21. The polypeptide of claim 20 wherein the polypeptide is provided on a polypeptide array.

22. A collection of polynucleotides, wherein the collection comprising the sequence information of at least one of SEQ ID NO:1-1786 and 3573-5358.

23. The collection of claim 22, wherein the collection is provided on a nucleic acid array.

24. The collection of claim 23, wherein the array detects full-matches to any one of the polynucleotides in the collection.

25. The collection of claim 23, wherein the array detects mismatches to any one of the polynucleotides in the collection.

26. The collection of claim 22, wherein the collection is provided in a computer-readable format.

27. A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising a polypeptide of claim 10 or 20 and a pharmaceutically acceptable carrier.

28. A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising an antibody that specifically binds to a polypeptide of claim 10 or 20 and a pharmaceutically acceptable carrier.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/34263

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C07H 21/04; C12N 15/11, 15/63, 15/70, 15/82, 15/85; C07K 14/00
US CL : 536/23.1; 435/320.1, 455, 468, 530/300, 350

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 536/23.1; 435/320.1, 455, 468, 530/300, 350

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
MEDLINE, EAST

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WAJIMA et al. The cDNA cloning and transient expression of an ovary-specific 17beta-hydroxysteroid dehydrogenase of chickens. Gene. 1999, Vol.233, pages 75-82	1-11, 13-16, and 19-26
A	US 5,175,095 A (MARTINEAU et al) 29 December 1992 (29.12.1992), see especially columns 3-18.	1-11, 13-16, and 19-26
A	Database PubMed, ID No. 2393392, FREUDENSTEIN et al. mRNA of bovine tissue inhibitor of metalloproteinase: sequence and expression in bovine ovarian tissue. Biochem. Biophys. Res. Commun. August 1990. Vol.171. No. 1. pages 250-256, see Abstract.	1-11, 13-16, and 19-26
A,P	Database PubMed, ID No. 10919256, HENNEBOLD et al. Ovary-selective genes I: the generation and characterization of an ovary-selective complementary deoxyribonucleic acid library. Endocrinology. August 2000. Vol.141. No.8. pages 2725-2734, see Abstract.	1-11, 13-16, and 19-26
A	Database PubMed, ID No. 2760883, BEIL et al. Synthesis of polypeptides by the cervix of the baboon (Papio anubis). J. Reprod. Fertil. July 1989. Vol.86. No.2. pages 535-544, see Abstract.	1-11, 13-16, and 19-26
A,P	Database PubMed, ID No. 10830289, HINSHELWOOD et al. A 278 bp region just upstream of the human CYP19 (aromatase) gene mediates ovary-specific expression in transgenic mice. Endocrinology. June 2000. Vol.141. No.6. pages 2050-2053, see Abstract.	1-11, 13-16, and 19-26

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
	07 JUN 2001
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230	Authorized officer Michael Woodward Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/34263

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
This includes 4 invention Groups and 3572 sequence species

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US00/34263

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional serch fees must be paid.
Group I, claims 1-11, 13-16, and 19-26, drawn to nucleic acid molecules, vector molecules and host cells containing said nucleic acids, polypeptides, methods of making said polypeptides and method of detection using said nucleic acids and polypeptides.
Group II, claim 12 and 28, drawn to antibodies and method of treatment using composition comprising said antibodies.
Group III, claims 17-18, drawn to methods of indentifying a binding partner to a polypeptides.
Group IV, claim 27, drawn to method of treatment using composition comprising polypeptides.

The inventions listed as Groups I-IV do not relate to a single inventive concept under PCT Rule 13.1 because, udner PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Group I encompasses nucleic acids, polypeptides expressed thereby, vectors and host cells containg same, respectively, and methods of making as well as the first method of use of this jubject matter. Groups II-V all are directed to different special technical features as summarized as follows: Group II is directed to an antibody and method of treatment using same, which antibody undergoes recognition and binding reactions wherein what is bound is different from what is bound by the compositions of Group I. For example, the polypeptides of Group I do not bind the polypeptides of Group I as the antibody of Group II does. Identification of binding partner and treatment are clearly different special technical features from detection. Group III is directed to the identification of a binding partner of a polypeptide, which is not identified in any of the other Groups and thus clearly contains its own special technical feature. Group IV is directed to treatment, which is a clearly different methods than the methods in the other Groups.Thus, in summary, each of Groups I-IV are directed to different special technical features and thus support this lack of unity.

Additionally, each of the claims is directed to more than one species of the generic invention. These species are deemed to lack unity of invention because they are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for more than one species to be searched, the appropriate additional search fees must be paid. The species are as follows: The claims include a series of polynucleotides and the polypeptides encoded thereby as representde by the sequences of SEQ ID Nos: 1-1786, and 3573-5358. Each of these polynucleotide sequences encodes a separate polypeptide and thus represent a separate gene. Therefore, each of these genes defines its own special technical feature. In summary, one species is a gene represented by one polynucleotide sequence and one polypeptide sequence encoded thereby.